DOCUMENT RESUME

ED 250 370 TM 840 684

AUTHOR Schmitt, Alicia P.; Crocker, Linda
TITLE The Relationship between Test Anxiety and Person Fit Measures.
PUB DATE Apr 84
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984). Small print in tables.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Ability; Achievement Tests; Attention; *Goodness of Fit; Junior High Schools; Latent Trait Theory; *Measurement Techniques; Multiple Regression Analysis; Response Style (Tests); *Test Anxiety; Test Norms
IDENTIFIERS Metropolitan Achievement Tests; *Person Fit Measures; Test Anxiety Scale for Adolescents

ABSTRACT

This study investigates the relationship between an examinee's test anxiety level, degree of person fit, and ability. Three major types of person fit indices measure the degree of unusual response patterns: norm comparison indices; goodness-of-fit indices; and extended indices. Five different person fit indices were calculated for the results of the Metropolitan Achievement Test (MAT) reading, mathematics, and science subscales of 413 seventh and eighth grade students: (1) the modified caution index; (2) the personal biserial correlation; (3) the norm-conformity index; (4) the Rasch person fit statistic; and (5) the extended caution index. The Test Anxiety Scale for Adolescents (TASA) was administered approximately two weeks prior to the MAT. Data were analyzed by a linear multiple regression analysis. High intercorrelations among the person fit indices confirmed previous finding. Significant interactions between ability and test anxiety demonstrated the moderating effect of ability levels on the person fit measures and test anxiety relationship. There was a direct relationship between person fit measures and test anxiety for lower ability examinees and an increasing negative relationship for high ability examinees. For examinees in the average ability ranges, there is little or no relationship between test anxiety and person misfit. The results are interpreted using the Cognitive-Attentional Theory of Test Anxiety. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED250370

THE RELATIONSHIP BETWEEN TEST ANXIETY AND PERSON FIT MEASURES

Alicia P. Schmitt
Office of Instructional Resources

Linda Crocker[1]
Foundations of Education

University of Florida

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Schmitt
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 684

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, April 1984.

[1]The authors are grateful to James Algina for helpful suggestions on analyses of the data in this study.

Test anxiety research findings present a consistent moderate negative correlation between test anxiety and total score measures of achievement (Tryon, 1980). High-test-anxious individuals tend to score lower on classroom and aptitude tests (Alpert and Haber, 1960; Harper, 1974; Mandler and Sarason, 1952; I. Sarason, 1975, 1963; Spielberger, et al., 1978). According to the Cognitive - Attentional Theory of Test Anxiety (CATTA), introduced by I. Sarason (1960) and extended by Wine (1971, 1981), the "major cognitive characteristics of test-anxious persons are negative self-preoccupation, and attention to evaluative cues to the detriment of test cues" (Wine, 1981, p. 137). Misdirection of attention, both in the pre-stages of evaluation (study phase) and the test taking situation, may limit coding, retention, and retrieval of information by high-test-anxious individuals. Thus, according to CATTA, performance of test-anxious persons will be negatively affected. To date most research on test anxiety has considered primarily its effects on examinees total test score.

Recently, Frary (1982) and Harnisch and Linn (1981) have suggested that test anxiety may be a factor which contributes to erratic performance of an examinee within a given test (i.e., missing relatively easy items, while answering more difficult items correctly). Careless errors and lack of concentration by high-test-anxious individuals could change the pattern of item responses from the pattern that would be expected. This would make the examinee appear to be an aberrant performer who "behaves as if his or her abilities were changing throughout the test" (Levine and Rubin, 1979, p. 272). To provide information about the consistency of examinee's response patterns, a variety of statistical procedures have been suggested which are generally called person fit indices.

These statistics indicate the deviation of the examinee response patterns from the norm response patterns (Frary and Giles, 1980; Frary, 1982; Harnisch and Linn, 1981; Levine and Rubin, 1979; Rudner, 1983; Tatsuoka and Tatsuoka, 1980; Tatsuoka and Linn, 1981). The purpose of this study was to investigate the relationship between an examinee's level of test anxiety and degree of person fit and to establish if this relationship is dependent on ability level.

The examinee group of interest in this study consisted of seventh and eighth graders responding to the Reading, Math and Science sub-scales of the Metropolitan Achievement Test (MAT). Five person fit indices were calculated. These were the modified caution index (MCI), the personal biserial correlation (PB), the norm-conformity index (NCI), the Rasch - X^2 statistic (RX^2) and the extended caution index (ECI).

PERSON FIT INDICES

Indices measuring the degree of unusual response patterns can be categorized into three major types: norm comparison indices, goodness-of-fit indices, and extended indices. Norm comparison indices, which are based on patterns of right and wrong answers, include: Sato's "caution-index", the "modified caution index" (Harnisch and Linn, 1981), the "agreement","disagreement", and "dependability" indices proposed by Kane and Brennan (1980), and the "norm-conformity index" (Tatsuoka and Tatsuoka, 1980). Goodness-of-fit or "appropriateness" indices are based on item response theory. Levine and Rubin (1979) and Rudner (1983) describe several indicators of the appropriateness of a score as an ability measure based on Rasch's one parameter logistic model and on Birnbaum's three parameter logistic model. Extended indices have been proposed by Tatsuoka and Linn (1981) as a link between the norm comparison and the goodness-of-fit indices.

The indices of person fit used in this study represent all three of these categories. The "modified caution index" (MCI), "personal biserial" (PB), and "norm-conformity index" (NCI) represent norm comparison indices. They are calculated by using test information organized in a Student-Problem (S-P) Table. The S-P Table is a matrix of zeros and ones, where the rows represent the students ranked from highest to lowest according to total test score. The columns represent the items arranged from left to right in ascending order of difficulty. Correct responses are represented by ones and incorrect responses are represented by zeros. Assuming that items are arranged in increasing order of difficulty (from easy to hard), a concordant response pattern is one in which an examinee answers the items correctly until he or she reaches an item that is too difficult and answers the items incorrectly from then on. The matrix would then have all ones in the upper left-hand corner, and all zeros in the lower right-hand corner. Some disruption of this pattern is expected, but when responses are affected by something other than ability, eratic response patterns develop. Norm comparison indices indicate the degree of aberrance for each student.

Harnisch and Linn (1981) introduced the modified caution index as a modification of an earlier caution index proposed by Sato (1975). The MCI has a lower bound of 0 and an upper bound of 1. The higher the value of the index, the more divergent is the person's response pattern. The formula for computing this index is presented in Appendix A.

The personal biserial correlation was proposed by Donlon and Fischer (1968). See Appendix A(2) for the formula for this index. The coefficient obtained represents the correlation between a person's item responses and the items difficulty values. Donlon and Fischer define item difficulty as the proportion of examinees who respond incorrectly to an item. Large values correspond to difficult items and small values correspond to easy items. A positive

correlation represents good fit, indicating that a person tends to answer correctly items that are easy for the group and miss the more difficult items. Low or negative correlations represent more divergent response patterns.

The Norm-Conformity Index (NCI) was developed by Tatsuoka and Tatsuoka (1980). The NCI indicates the degree of concordance to a group response pattern where items are arranged in descending order of difficulty, from hardest to easiest. Values of this index may range from -1 to 1. The smaller the index, the more divergent is the individuals' response pattern in comparison to the group norm. Let S denote the row vector of a person's response pattern; let $\bar{S}$ denote the transpose of the complement of S; and let $N=\bar{S}'S$. The formula for the NCI is presented as (3) in Appendix A. This index is undefined for either perfect or zero scores.

Rasch Person Fit Statistic (RX^2), also referred to as Maximum Likelihood Procedure (MAX) (Wright and Panchapakesan, 1969) and as a weighted total fit mean square (Rudner, 1983), was adopted for use with the one parameter Rasch model under the BICAL program (Wright, et al., 1979). This index belongs to the goodness-of-fit category of person fit indices and is based on item response theory. It has a high value for an examinee who has a response pattern that is inconsistent with the examinee's score and the Rasch model measure of item difficulty. The formula is shown as (4) in Appendix A.

Tatsuoka and Linn (1981) described the extended caution index as a linear transformation of the distance between a person's response pattern and a theoretical curve. In the case of ECI, this curve is the Group Response Curve (GRC), which is "an average function of N different Person Response Curves" (Tatsuoka and Linn, 1981, p. 10). The caution index is an extension of norm comparison indices to the goodness of fit indices. For the ECI, probabilities of success, calculated through item response theory-logistic models substitute the zeros or ones in the S-P Table. For the purposes

of this study, the one-parameter logistic model (Rasch) will be used to calculate these probabilities. The formula is presented as (5) in Appendix A. This formula is computed by the ratio of two covariances. The higher the value of the ECI, the more variation from the expected response pattern. This index is also limited by perfect or zero scores. The denominator would become zero and the value infinite.

Although these various person-fit indices have usually been found to be highly correlated for the same test (e.g., Harnisch and Linn, 1981), it seemed possible that they would not be equally correlated to levels of test anxiety, particularly when ability-level was taken into account. There was, for example, some expectation that stronger relationships might be detected between test anxiety and the norm comparison indices, such as the personal biserial, rather than for goodness-of-fit indices such as the Rasch, simply because the standardized tests used in this study might not meet the fairly strong assumptions required for these models. Furthermore, Rudner (1983) has reported that the Rasch person-fit statistic is more sensitive to aberrant responses at higher ability levels; from test anxiety theory, we might therefore suspect that examinees at higher ability levels (who are low in test anxiety) would have lower misfit than high ability ability examinees who are high in test anxiety.

METHODOLOGY

Sample

The participants in this study were 225 seventh graders, and 188 eighth graders from a metropolitan middle school with varying socioeconomic levels. An almost even distribution of each sex was found at each grade level. Approximately 70% of the students were white and 30% were black at each grade level.

Instruments

The Test Anxiety Scale for Adolescents (TASA) (Schmitt and Crocker, 1982) was used to measure test anxiety. This scale consists of 31 true-false items and is designed for use with examinees in middle school or junior high grades. TASA is adapted from the Mandler Sarason Test Anxiety Scale which is designed for use with adults. Schmitt and Crocker (1982) found the factor structure of the TASA to be fairly similar to the adult test anxiety scale. They reported a KR20 of .87 as a total score reliability estimate for middle school examinees. Some evidence of the instrument's construct validity was reported by Simmons (1983) who found moderate negative correlations with performance of fifth graders on a state minimum competency test. All items deal exclusively with examinee feelings about tests. Sample items include:

"I worry just before getting a test back"; and

"Sometimes on a test I just can't think."

Data Collection Procedure

The Metropolitan Achievement Test subscales (Form KS) were used to measure achievement in reading, mathematics, and science, (Prescott, et al., 1978). The TASA was administered during the spring semester, approximately two weeks prior to the administration of the MAT. Homeroom teachers administered all tests during the normal first period class. The MAT was administered as part of the regular school testing program.

Data Analysis

To investigate the relationship between examinees' level of test anxiety and degree of person-fit and to study if this relationship is dependent upon ability level, a linear multiple regression analysis was used. In this analysis person-fit measures were the dependent variables and ability (reading, math or science) and TASA were the continuous independent variables. The model used for each ability measure and person-fit index is:

$$Y' = b_o + b_1X_1 + b_2X_2 + b_3X_1X_2$$

where

Y' = person fit predicted by the model,

b_o = intercept value,

b_1 = regression slope for each independent variable,

X_1 = ability, estimated from total score,

X_2 = TASA,

X_1X_2 = interaction between ability and TASA.

RESULTS

The mean, standard deviation, and minimum and maximum values for person fit measures, TASA, and ability scores by grade and subject content area are presented in Table 1.

Insert Table 1 About Here

Person-fit means and standard deviations by index are very similar across subject content areas and grades. Person-fit measure intercorrelations for each grade level are presented in Table 2. These results indicate that person-

fit measures are highly related within same subject content area. These correlations range from .78 to .99. The Rasch chi square (RX^2) index has the lowest relationships to other indices within same subject content areas. Interestingly RX^2 is the only index with significant intercorrelations with other indices across subject content areas.

Insert Table 2 About Here

Correlations between person-fit measures, ability measures of reading, math and science and TASA for each grade are presented in Table 3. Correlation values between person-fit measures and TASA range from .02 to .22. The highest correlation between person-fit measures and TASA occurred for seventh grade science person-fit values and for eight grade reading person-fit values. Correlations between person-fit measures and their corresponding total ability scores range from .00 to .50. The Personal Biserial (PB) index seems to have the lowest correlation with total score.

Insert Table 3 About Here

Multiple regression results for the model with ability, TASA, and their interaction are summarized in Table 4. Values of R^2 for the model and those with significant interactions are identified.

Insert Table 4 About Here

For the seventh grade sample, a significant proportion of variance in the modified caution index was explained in each of the three subject area tests by the linear combination of test anxiety, ability and their interaction. Only in reading, however, was the percentage of variance explained greater than 10%, and on this test the interaction of ability and test anxiety was significant. For the personal biserial, the norm-conformity index and the extended caution index, although several significant R^2 values were observed, these were all less than .10. Thus it is difficult to interpret these relationships as being substantially important. For the Rasch χ^2 index of person fit, substantial proportions of variation were explained by the model in the areas of math and science. No significant interaction effect of ability and test anxiety was detected in either of these cases.

For the eighth grade, the modified caution index again appears to be sensitive to variation in examinees' levels of test anxiety, ability and their interaction. The significant R^2 values exceeded 10% for this index in reading and science, and for science, with the largest R^2(.23), significant interaction of test anxiety and ability was detected. For the personal biserial, the nonconformity index and the extended caution index, the significant R^2 values greater than .10 were found only in the area of science, and for each of these cases, the interaction between ability and test anxiety contributed significantly to the overall model. Significant (and meaningful) R^2 values for the Rasch index were found for the areas of science and mathematics without any interaction between ability and test anxiety.

For further interpretations of these results, two major questions were addressed:

1. What is the nature of the interaction effect of test anxiety and ability on person fit?

2. When no interaction occurred, what are the contributions of ability and test anxiety to the variations in person-fit?

Significant interactions were followed up by graphing the relationship between person fit measures and TASA at selected ability levels. The formula to calculate the regression line at each ability level is:

$$Y' = b_0 + b_1X_1 + (b_2 + b_3X_1)X_2$$

At any value of ability (X_1) a predicted person fit measure (Y') was calculated for different points of TASA (X_2). The intercept of this equals ($b_0 + b_1X_1$) and the slope equals ($b_2 + b_3X_1$). Table 5 shows the slope and intercept estimates which were used in plotting these lines for all cases which R^2 exceeded .10.

Insert Table 5 About Here

The regression lines resulting from these computations are shown in Figures 1-5. From these graphs we can clearly see the nature of the interaction of ability and test anxiety on various indices of person fit. Although the points of intersection of these lines may vary, the same general pattern of this relationship occurs in all cases. Namely, this pattern of interaction can be characterized as follows:

1. For examinees in the average ability ranges, there is little or no relationship between test anxiety and person misfit. (Note the "flat" slope of the regression line for Group E in all figures.)
2. As examinee ability level increases (see lines for Groups F, G, and H), the slope of the regression lines increases, generally indicating an increasing negative relationship between test anxiety and person fit. Specifically high-ability, low-anxious examinees show more misfit than high-ability, high-anxious examinees.
3. As examinee ability decreases, (see lines for Groups C, B, and A), the opposite effect occurs. Namely low-ability, low-anxious examinees show less misfit than low-ability, high-anxious examinees.

Insert Figures 1-5 About Here

When no interaction occured, only ability main effects were significant. Table 6 presents the type I sums of squares which measures incremental sums of squares for the model as each variable is added.

Insert Table 6 About Here

DISCUSSION

High intercorrelations among the person fit indices in this study confirm previous findings (Harnish and Linn, 1981 and Rudner, 1983). Only multiple regression models accounting for at least 10% of the variance were considered in the interpretation of results. Because of the large sample size a number of fairly small R^2 were statistically significant, so only cases where there was a meaningful percentage of variance were considered.

Significant interactions between ability and test anxiety demonstrate that ability levels moderate the relationship between person fit measures and test anxiety. For lower ability examinees there is a direct relationship between person fit measures and test anxiety. For high ability examinees this relationship is inverse.

Figures 1 and 2 present the two general pictures of these interactions. Figure 1 shows the relationship between the Modified Caution Index and test anxiety as measured by TASA for Seventh grade examinees with different reading levels. The nature of the interaction is the same as previously described; but higher ability level students have higher overall MCI values (more misfit). For all the other significant interactions which were for person fit measures of science at the eight grade, lower ability level students have higher values of misfit. Figure 2 shows this for the MCI. The difference found for these two content areas might be due to the performance of this

sample on the reading subtest. More than 10% of the seventh graders missed the last fifteen items of the reading test; making it appear more like a speeded test than a power test. Higher ability examinees probably attempted more end items, increasing their probability of getting higher misfit classifications. A study of speeded, power tests, and person fit measurement is currently under way by these authors.

The Cognitive-Attentional Theory of Test Anxiety and previous person fit research findings suggest interpretive explanations of the interaction results. According to Tobias (1980) and Weinstein, et. al. (1982) high-test-anxious students will have worse performance on difficult materials compared to low-test-anxious individuals, while with easier material little difference between anxiety levels is expected. In reporting results about cognitive coping behavior and anxiety, Houston (1982) suggests that "...highly trait-anxious (and test anxious) individuals tend to lack organized ways of coping with stress and instead ruminate about themselves and the situation in which they find themselves" (p. 198).

If high-test anxious students' performance on more difficult items is more affected by their anxiety as these examinees reach items that have a level of difficulty that approximates their ability level they will have a harder time coping. Other testing strategies such as test wiseness would not be readily available due to lack of concentration. Examinees with high ability levels and low test anxiety could take advantage of test-taking skills and answer item correctly beyond their ability level. Since these items would not be answered with the same degree of certainty as easier items, more deviation from the expected response pattern could occur and higher misfit would result. For examinees with high ability and high test anxiety, coping and test-taking skills could be blocked; making them consistently miss harder items. Lower misfit values would occur for this group. Examinees with lower

ability and lower test anxiety levels would answer items correctly to the point where they reach items at their ability level and then miss the harder items. Some misfit could occur due to attempts at harder items. For examinees with low ability and high test anxiety, distracting thoughts might interfere with performance to almost all items. Even easy items (in reference to their ability level) could be missed; this would classify this group as high misfits. A follow-up comparison of the response patterns for these groups and their correspondence to the predicted patterns is planned.

These interaction effects between ability and test anxiety seem to appear more consistently for science especially at the eighth grade. One possible explanation is that the standardized science test fits the curriculum less than the reading or math tests. The mean item difficulty for the science test is lower, indicating a harder test. Examinees taking the science examination might find themselves in a more ambiguous and hence anxiety-producing situation.

Regression models with only significant main effects of ability were found mainly for $R\chi^2$ measures of person fit. These results support findings of Rudner (1983) and may suggest that the $R\chi^2$ index might be more sensitive to overall ability level than to anything else. These findings also support speculations by Frary (1982) and Harnisch and Linn (1981) that test anxiety can contribute to erratic performance of an examinee. High ability students can be quite anxious thereby affecting their performance and response pattern. Results of this study extend the interpretive value of misfit statistics. Attempts to understand why an examinee "misfit" a test should include consideration of the examinee's level of ability and test anxiety. For text anxiety construct development the relationships between person misfit and test anxiety endorse the CATTA. More research comparing the different indices and using other samples in the study of these relationships is needed.

REFERENCES

Alpert, R., and Haber, R.N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Donlon, T.F., and Fischer, F.E. An index of an individual's agreement with group-determined item difficulties. Educational and Psychological Measurement, 1968, 28, 105-113.

Frary, R.B. A comparison among person-fit measures. Paper presented at the annual meeting of the American Educational Research Association, New York, 1982.

Frary, R.B. and Giles, M.B. Multiple-choice test bias due to answering strategy variation. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, Massachusetts, 1980.

Harnisch, D.L., and Linn, R.L. Analysis of item response patterns: Questionable test data and dissimilar curricula practices. Journal of Educational Measurement, 1981 18, 133-146.

Harper, F.B.W. The comparative validity of the Mandler-Sarason Test Anxiety Questionnaire and the Achievement Anxiety Test. Educational and Psychological Measurement, 1974, 34, 961-966.

Houston, K. Trait anxiety and cognitive coping behavior. In I.G. Sarason and C.D. Spielberger (Eds.), Achievement, Stress, and Anxiety. Washington, DC: Hemisphere, 1982.

Levine, M.V., and Rubin, D.B. Measuring the appropriateness of multiple choice test scores. Journal of Educational Statistics, 1979, 4, 269-290.

Kane, M.T., and Brennan, R.L. Agreement coefficients as indices of dependability for domain-referenced tests. Applied Psychological Measurement, 1980, 4, 105-126..

Mandler, G., and Sarason, S.B. A study of anxiety and learning. Journal of Abnormal and Social Psychology, 1952, 47, 166-173.

Prescott, G.A., Balow, I.H., Hogan, T.P., and Farr, R.C. Teacher's manual for administering and interpreting-Metropolitan Achievement Tests (Advanced 1, Forms JS and KS). The Psychological Corporation, 1978.

Rudner, L.M. Individual assessment accuracy. Journal of Educational Measurement, 1983, 20, 207-220.

Sarason, I.G. Empirical findings and theoretical problems in the use of anxiety scales. Psychological Bulletin, 1960, 57, 403-415.

Sarason, I.G. Test anxiety and intellectual performance. Journal of Abnormal and Social Psychology, 1963, 66, 73-75.

Sarason, I.G. Anxiety and self-preoccupation. In I.G. Sarason and E.D. Spielberger (Eds.), Stress and Anxiety (Vol. 2). Washington, DC: Hemisphere, 1975.

Sato, T. The construction and interpretation of S-P Tables. Tokyo: Meiji Tosho, 1975 (in Japanese; cited in Harnisch and Linn, 1981).

Schmitt, A.P., and Crocker, L. Test anxiety and its components for middle school students. Journal of Early Adolescence, 1982, 2, 267-275.

Spielberger, C.D., Gonzalez, H.P., Taylor, C.J., Algaze, B., and Anton, W.D. Examination stress and test anxiety. In C.D. Spielberger and I.G. Sarason (Eds.), Stress and anxiety (vol. 5). New York: Hemisphere/Wiley, 1978.

Tatsuoka, K., and Linn, R.L. Indices for detecting unusual item response patterns in personnel testing: Links between Direct and Item-response-theory approaches. (Research Report 81-5) Urbana, Il.: University of Illinois, Computer-based Education Research Laboratory, 1981.

Tatsuoka, K., and Tatsuoka, M.M. Detection of aberrant response patterns and their effect on dimensionality. (Research Report 80-4) Urbana, Il.: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tobias, S. Anxiety and instruction. In I.G. Sarason (Ed.), Test Anxiety: Theory, Research, and Applications. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Tryon, G.S. The measurement and treatment of test anxiety. Review of Educational Research, 1980, 50, 343-372.

Weinstein, C.E., Cubberly, W.E., and Richardson, F.C. The effects of test anxiety on learning at superficial and deep levels of processing. Contemporary Educational Psychology, 1982, 7, 107-112.

Wine, J.D. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Wine, J.D. Cognitive-attentional theory of test anxiety. In I.G. Sarason (Ed.), Test Anxiety: Theory, Research, and Applications. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

Wright, B.D., Mead, R., and Bell, S. BICAL: Calibrating items with the Rasch model (RM-23b). University of Chicago, 1979.

Wright B.D., and Panchapakesan, N.A. A procedure for sample free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Table 1

Descriptive Statistics for Person Fit Measure,

Test Anxiety, and Ability Scores by Grade and Content Area

Person Fit Measure	Grade - 7 n	Mean	SD	Min.	Max.	Grade - 8 n	Mean	SD	Min.	Max.
Reading:	224					184				
MCI		.23	.10	.02	.81		.26	.12	0	.69
PB		.57	.20	-.46	1.12		.59	.23	-.12	1.12
NCI		.54	.20	-.62	.96		.55	.21	-.10	1.00
RX2		.98	.14	.65	1.37		.97	.15	.65	1.49
ECI		.01	.35	-.89	1.92		.01	.37	-.93	1.24
Math:	225					188				
MCI		.12	.08	.06	.50		.23	.09	.04	.58
PB		.57	.17	.02	.93		.58	.19	-.20	.98
NCI		.55	.17	.03	.89		.56	.18	-.17	.92
RX2		.98	.14	.67	1.43		.98	.15	.62	1.67
ECI		-.01	.31	-.62	.96		-.01	.33	-.72	1.34
Science:	225					186				
MCI		.25	.08	.07	.59		.26	.08	.07	.54
PB		.49	.16	-.14	.88		.49	.16	-.07	.78
NCI		.47	.17	-.16	.87		.47	.17	-.09	.86
RX2		.99	.12	.69	1.40		.99	.12	.78	1.48
ECI		-.04	.36	-.94	1.28		-.03	.38	-.90	1.13
Totals:										
Reading		35.39	11.62	9	54		39.47	10.48	15	54
Math		30.32	9.44	13	49		33.32	8.88	8	48
Science		33.43	9.87	9	53		35.02	9.13	14	52
TASA		15.05	6.66	1	29		13.75	6.23	1	27

BEST COPY AVAILABLE

Table 2

Person Fit Measures Intercorrelations by Subject Content Area

7th grade - Shown Above Diagonal, 8th grade - Below Diagonal

	Reading:					Math:					Science				
		7thm=224		8thm=184			7thm=223		8thm=188			7thm=225		8thm=186	
	MCI	PB	NCI	RX_2	ECI	MCI	PB	NCI	RX_2	ECI	MCI	PB	NCI	RX_2	ECI
Reading:															
MCI		-.97**	-.91**	.82**	.88**										
PB	-.86**		.94**	-.88**	-.99**										
NCI	-.81**	.92**		-.82**	-.96**										
RX_2	.78**	-.90**	.83**		.88**										
ECI	.93**	-.98**	-.93**	-.90**											
Math:															
MCI							-.98**	-.99**	.93**	.99**					
PB						-.98**		.98**	-.91**	-.97**					
NCI						-.99**	.97**		-.93**	-.99**					
RX_2						.89**	-.88**	-.91**		.94**					
ECI						.99**	-.97**	-.99**	.90**						
Science:															
MCI												-.98**	-.99**	.93**	.99**
PB											-.97**		.97**	-.90**	-.97**
NCI											-.99**	.97**		-.91**	-.99**
RX_2											.95**	-.94**	-.94**		.92**
ECI											.99**	-.96**	-.99**	.94**	

*Significant at alpha = .01

**Significant at alpha = .0001

—Not significant.

BEST COPY AVAILABLE

Table 3

Correlations Between Person Fit Measures, Reading, Math, Science, and Tasa

by grade

	GRADE - 7				GRADE - 8			
	Totals: N=224 READING	225 MATH	225 SCIENCE	237 TASA	Totals: N=184 READING	188 MATH	186 SCIENCE	198 TASA
Reading:								
MCI	.28**	.20**	.26**	-.18**	.32**	.23**	.31**	—
PB	—	—	—	—	.18*	—	—	-.15*
NCI	—	—	—	—	.20*	.15*	—	-.15*
$R\chi^2$	—	—	—	—	-.16*	—	—	.17*
ECI	—	—	—	—	—	—	—	—
Math:								
MCI	—	—	—	—	—	—	—	—
PB	—	—	—	—	—	—	—	—
NCI	—	.20**	—	—	—	.17*	—	—
$R\chi^2$	-.25**	-.32**	-.22*	—	—	-.30**	-.13*	—
ECI	-.18**	-.23**	-.15	—	—	-.28**	—	—
Science:								
MCI	-.20**	-.27**	-.25**	.17*	-.34**	-.32**	-.45**	—
PB	—	—	—	—	.25**	.22**	.29**	—
NCI	—	.25**	.22**	-.16*	.35**	.32**	.44**	—
$R\chi^2$	-.31**	-.39**	-.39**	.22**	-.42**	-.38**	-.50**	—
ECI	-.22**	-.29**	-.28**	.18**	-.39**	-.36**	-.30**	—
Totals:								
READING	—	.73**	.83**	-.38**	—	.70**	.79**	-.17*
MATH		—	.76**	-.44**	—	—	.75**	-.26**
SCIENCE	—	—	—	-.39**	—	—	—	-.27**
TASA	—	—	—	—	—	—	—	—

* Significant at alpha = .05
** Significant at alpha = .01

BEST COPY AVAILABLE

Table 4

Summary of Significant R^2 for Model with Ability TASA and their Interactions

Grade:	Ability	MCI	PS	MCI	RX^2	ECI
7th						
	Reading	.11**	.04*	.01	.02	.03
	Math	.04*	.03	.05*	.11**	.06**
	Science	.07**	.01	.06**	.16**	.09**
8th						
	Reading	.12**	.05*	.05*	.05*	.02
	Math	.04	.04	.06*	.10**	.07**
	Science	.23**	.13**	.23**	.28**	.10**

*Significant at alpha = .05

**Significant at alpha = .01

__Significant interaction

BEST COPY AVAILABLE

Table 5

Significant Ability * TASA Interaction Parameter Estimates and R^2 Increase-Interaction for Person Fit Measures

Dependent Variable	Parameter	Estimate	Standard Error	t	R^2 Increase
7th-Reading					
	Intercept	.029	.064	.46	
MCI	Slope-Ability	.006	.002	3.74**	
	Slope-TASA	.007	.004	2.02*	
	Slope-Ability*TASA	-.000	.000	-2.52*	.026
	(Model - R^2 = .11**; F3,220 = 8.90**)				
8th-Science					
	Intercept	.314	.059	5.28**	
	Slope-Ability	-.001	.002	-.86	
MCI	Slope-TASA	.006	.004	1.73	
	Slope-Ability*TASA	-.000	.000	-2.15*	.020
	(Model - R^2 = .23**; F3,182 = 18.00**)				
	Intercept	.532	.120	4.44**	
	Slope-Ability	-.002	.003	-.63	
PB	Slope-TASA	-.017	.007	-2.23*	
	Slope-Ability*TASA	.001	.000	2.66**	.034
	(Model - R^2 = .13**; F3,182 = 9.06**)				
	Intercept	.373	.121	3.08**	
	Slope-Ability	.002	.003	.66	
NCI	Slope-TASA	-.014	.008	-1.91	
	Slope-Ability*TASA	.000	.000	2.33*	.023
	(Model - R^2 = .23**; F3,182 = 17.93**)				
	Intercept	.229	.257	.89	
	Slope-Ability	-.006	.007	-.93	
ECI	Slope-TASA	.033	.016	2.04*	
	Slope-Ability*TASA	-.001	.000	-2.50*	.025
	(Model - R^2 = .28**; F3,182 = 23.56**)				

*Significant at alpha = .05

**Significant at alpha = .01

BEST COPY AVAILABLE

Table 6

Significant Main Effect of Ability

as Predictor of Person-Fit Measure

Dependent Variable	Source	df	Type I Sum Squares	Mean Squares	F	R^2
Grade 7: EX^2 (Math)	Modal	3	.50	.17	9.20**	.11**
	Math	1	.47	.47	16.00**	
	TASA	1	.01	.01	.47	
	Math*TASA	1	.02	.02	1.12	
	Error	221	4.01	.02		
Grade 7: EX^2 (Science)	Modal	3	.50	.17	13.73**	.16**
	Science	1	.49	.49	39.95**	
	TASA	1	.01	.01	1.21	
	Science*TASA	1	.00	.00	.01	
	Error	221	2.70	.01		
Grade 8: MXI (Reading)	Modal	3	.31	.10	7.97**	.12**
	Reading	1	.27	.27	20.38**	
	TASA	1	.04	.04	3.36	
	Reading*TASA	1	.00	.00	.17	
	Error	180	2.36	.01		
Grade 8: EX^2 (Math)	Modal	3	.44	.15	7.01**	.10**
	Math	1	.38	.38	18.06**	
	TASA	1	.01	.01	.25	
	Math*TASA	1	.06	.06	2.73	
	Error	184	3.86	.02		
Grade 8: EX^2 (Science)	Modal	3	.67	.22	22.30**	.27**
	Science	1	.63	.63	63.02**	
	TASA	1	.01	.01	.81	
	Science*TASA	1	.03	.03	3.07	
	Error	182	1.81	.01		

*Significant at alpha= .05
**Significant at alpha= .01

BEST COPY AVAILABLE

BEST COPY AVAILABLE

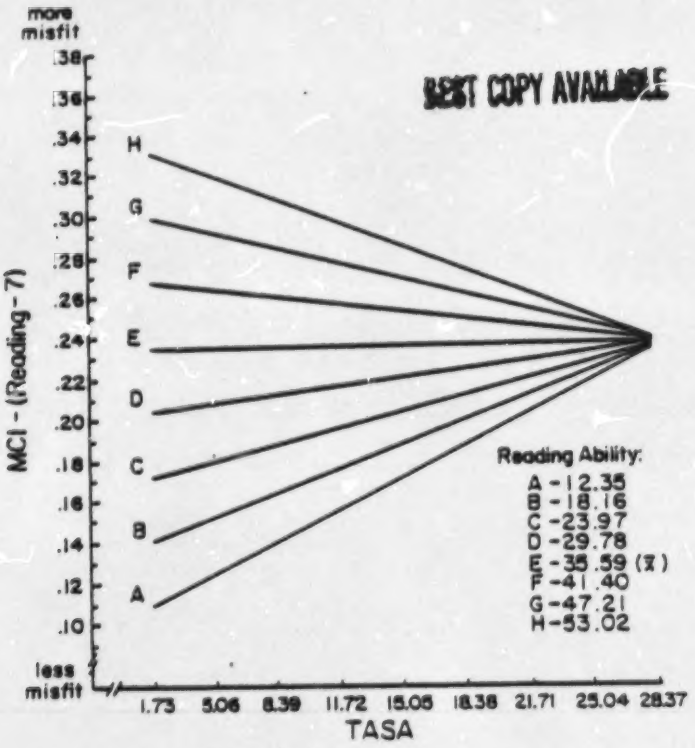

Figure 1. Relationship between the Modified Caution Index and TASA for 7th grade examinees at different reading ability levels.

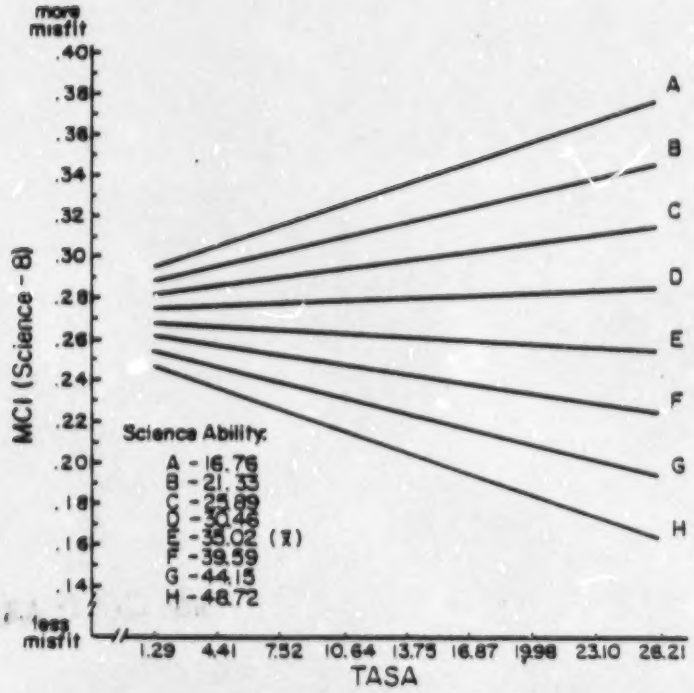

Figure 2. Relationship between the Modified Caution Index and TASA for 8th grade examinees at different science ability levels.

BEST COPY AVAILABLE

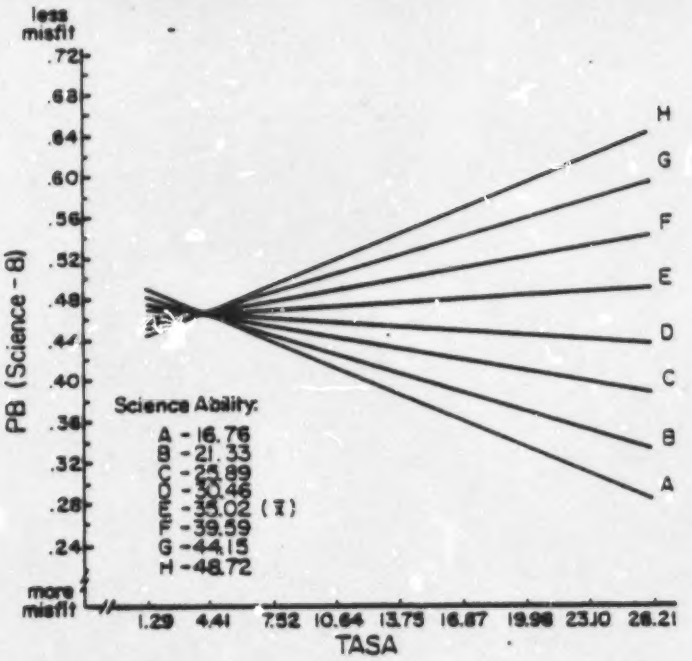

Figure 3. Relationship between the Personal Biserial and TASA for 8th grade examinees at different science ability levels.

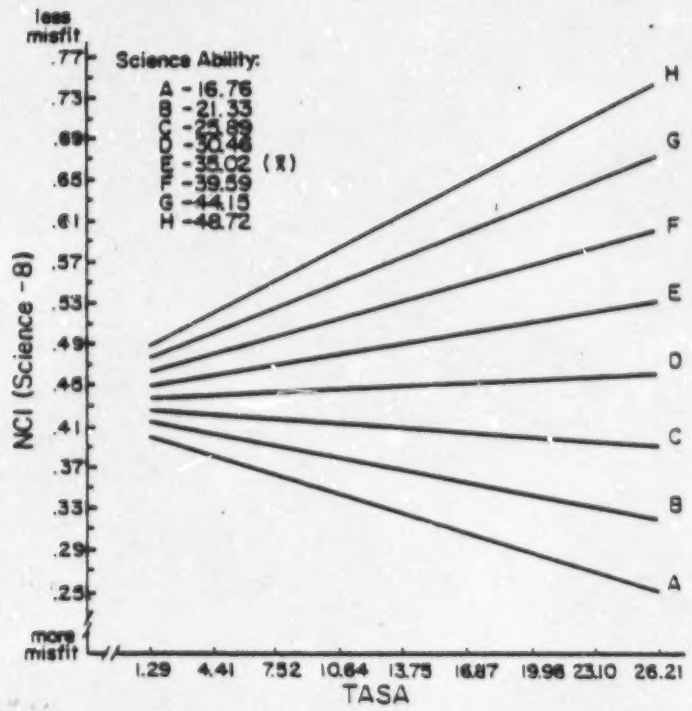

Figure 4. Relationship between the Norm Conformity Index and TASA for 8th grade examinees at different science ability levels.

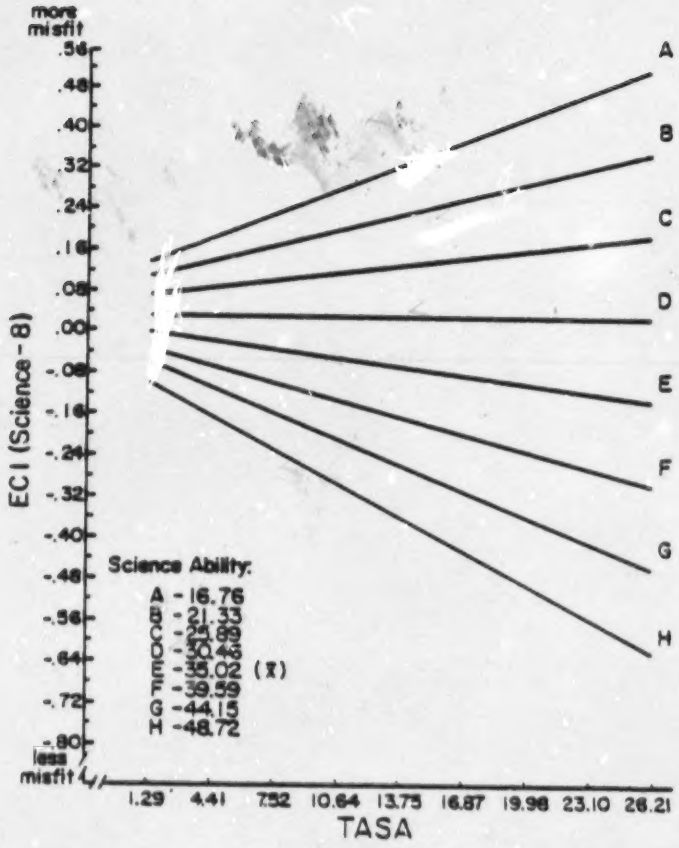

Figure 5. Relationship between the Extended Caution Index and TASA for 8th grade examinees at different science ability levels.

BEST COPY AVAILABLE

APPENDIX A

Person Fit Statistics

1. **Modified Caution Index:**

$$MCI = \frac{\sum_{j=1}^{n_{i.}} (1 - U_{ij}) n_{.j} - \sum_{j=(n_{i.}+1)}^{J} U_{ij} n_{.j}}{\sum_{j=1}^{n_i} n_{.j} - \sum_{j=(J^j+1-n_i)}^{J} n_j}$$

where i is the examinee index, j is the item index, U_{ij} is 1 if examinee i answers item j correctly and 0 if examinee i answers item j incorrectly, n_i is the total number of correct responses for examinee i, and n_j is the number of correct responses to item j.

2. **Personal Biserial:**

$$PB = \frac{(\bar{Q}_r - \bar{Q}_c)}{S_{Q_r}} \cdot \frac{J_r'}{Y}$$

where $\bar{Q}_r$ is the mean item difficulty for items answered, $\bar{Q}_c$ is the mean item difficulty for items answered correctly, S_{Q_r} is the standard deviation for Q_r, J_r' is the number of items answered correctly divided by the number of items answered, and Y is the ordinate from the standard normal curve at the point separating the proportions J_r' and $(1-J_r')$.

3. **Norm Conformity Index:**

$$NCI = \frac{2U_a}{U-1}$$

where $U_a = \sum_{ij>1} \sum n_{ij}$, the sum of the above diagonal elements of N and

$U = \Sigma n_{ij}$, the sum of all the elements of N.

4. **The Rasch Chi Squared Index:**

$$R\chi^2 = \frac{\sum_{j=1}^{J} (U_{ij} - P_{ij})^2}{\sum_{j=1}^{J} P_{ij}(1-P_{ij})}$$

where U_{ij} is the response of examinee i to item j and P_{ij} is the probability of a correct response for examinee i on item j as predicted by the Rasch model.

$$P_{ij} = e^{(\theta_i - b_j)}/[1 + e^{(\theta_i - b_j)}]$$

5. Extended Caution Index:

$$ECI = 1 - \frac{\sum_{j=1}^{J} (Y_{ij} - P_{i.})(Y_{.j} - P_{..})}{\sum_{j=1}^{J} (\hat{P}_{ij} - \hat{P}_{ij}/J)(Y_{.j} - P_{..})}$$

where Y_{ij} is the response of examinee i to item j, $Y_{.j}$ is the sum across examinees for item j, $P_{i.}$ is the proportion of correct responses of examinee i, $P_{..}$ is the total proportion of correct responses, $\hat{P}_{ij}$ is the probability of correct response for each examinee i on item j according to the Rasch model, and $\Sigma\hat{P}_{ij}/J$ is the test ability parameter for each examinee i.